Regioselectivity and Chemoselectivity Analysis of Oxazole and Thiazole Ring Formation by the Peptide-Heterocyclizing Microcin B17 Synthetase Using High-Resolution MS/MS

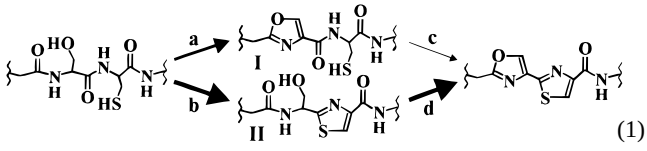
Neil L. Kelleher, Peter J. Belshaw, and Christopher T. Walsh*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School Boston, Massachusetts 02115

Received June 25, 1998

The presence of rigid structural elements such as thiazole and oxazole heterocyclic rings in peptide-derived compounds confers a wide range of therapeutic properties including antibiotic, antiviral, and antitumor activity. Molecules with thiazoles and oxazoles in tandem 4,2-connected bisheterocyclic arrays are of particular interest, given the recent demonstration that the bisthiazole moiety in the antitumor drug bleomycin is a DNA-targeting intercalator.[1] The tandem oxazole/thiazole and thiazole/oxazole pairs found in the *Escherichia coli* antibiotic microcin B17 (MccB17)[2] may target MccB17 to DNA and kill bacteria by the accumulation of double strand DNA breaks in a DNA gyrase dependent mechanism.[3] Posttranslational ring biogenesis from Ser and Cys residues occurs on a complex of three proteins (McbB, C, D), utilizing Zn^{2+}, FMN, and ATP to cyclize, dehydrate, and aromatize these substrate residues.[4] Using high-resolution tandem mass spectrometry (MS/MS), we determine here the effect of heterocycle formation on MS/MS of reaction intermediates and use this unique MS/MS signature to ascertain the regio- and chemoselectivity of MccB17 synthetase.

The first heterocyclizable site in the 69 residue McbA substrate for posttranslational processing by MccB17 synthetase is the tripeptidyl unit $Gly_{39}Ser_{40}Cys_{41}$, converted to the tandem bisheterocycle.[5a,b] Two rounds of cyclodehydration (−18.01 Da) and dehydrogenation (−2.02 Da) occur (eq 1) in both full length



(1)

substrate ($McbA_{1-69}$) and a $McbA_{1-46}$ fragment. Discrete M-20 Da species accumulate in enzyme incubations, which then are converted to bisheterocycle.[5b] These intermediates are presumed to be monocyclic species, and characterization would allow interrogation of preference of the bisheterocyclizing synthetase for regioselectivity and/or chemoselectivity in ring formation. McbA substrate recognition requirements[5a,6] mandate that mechanistic investigations be conducted in at least the $McbA_{1-46}$ substrate context (4.3 kDa). Substrates were generated as MBP−$McbA_{1-46}$ fusion proteins,[5] subjected to MccB17 synthetase, and then cleaved at a thrombin site to release 48mer fragments GS−$McbA_{1-46}$ (see Figure 1c) which were transferred into the gas phase by electrospray ionization (ESI)[7] and analyzed in a 9.4[8a] or 4.7[8b] Tesla Fourier transform (FT) mass spectrometer.[9]

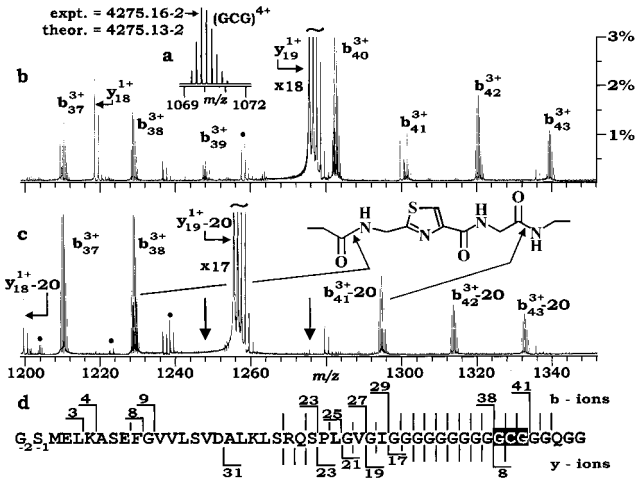


Figure 1. (a) Partial ESI/FT mass spectrum of the $Gly_{39}Cys_{40}Gly_{41}$ substrate (48 residues) of microcin B17 synthetase; 4+ ions, 10 scans, 9.4 T. (b) Partial MS/MS spectrum (*m*/*z* 1200−1350) from photodissociation of the $Gly_{39}Cys_{40}Gly_{41}$ substrate ions in (a); dot, H_2O loss peak; 500 scans, 9.4 T. (c) Partial MS/MS spectrum from photodissociation of the 4+ ions of the Gly thiazole−Gly enzymatic product; vertical arrows, theoretical positions of b_{39}^{3+} and $(b_{40}\text{-}20)^{3+}$ fragment ions; dots, H_2O loss peaks; 500 scans, 9.4 T. (d) Correlation of all fragment ions observed in the full MS/MS spectrum of the $(GCG)^{4+}$ ions with the primary sequence of this substrate; b- and y-type ions are indicated with vertical lines above and below the sequence, respectively.

For MS/MS, it was anticipated that starting substrates could be sequenced partially via b-type (containing the N-terminus) and y-type (containing the C-terminus)[10] fragment ions and that conversion of a peptide linkage into a thiazole or oxazole should abrogate cleavage at that site. As an initial test, a substrate mutant with only a single cyclizable residue, MBP−$McbA_{1-46}$ $G_{39}C_{40}G_{41}$ (GCG), yielded multiply charged species with 3+, 4+, and 5+ charge states and the expected relative molecular weight (M_r) value (4275.16-*2* vs 4275.13-*2* Da,[11] theory; 7 ppm error). The $(GCG)^{4+}$ ions (Figure 1a) were of highest abundance for subsequent isolation[12] and fragmentation,[13] producing 20 y-type and 28 b-type fragment ions (Figure 1d) in the MS/MS spectrum. The b-ion series (b_{37}^{3+}−b_{43}^{3+}) between *m*/*z* 1200−1350 (Figure 1b) directly sequences the predicted site of thiazole formation. Conversion of $G_{39}C_{40}G_{41}$ to the amidomethylthiazole-G_{41} product results in a loss of 20.02 Da ($\Delta m_{theory} = -20.03$ Da, data not shown); MS/MS of these $(GCG\text{-}20)^{4+}$ ions reveals an absence of both the b_{39}^{3+} and the $(b_{40}\text{-}20)^{3+}$ fragment ions (Figure 1c, vertical

* Corresponding author: Fax: (617) 432-0556. E-mail: walsh@walsh.med.harvard.edu.

(1) Wu, W.; Vanderwall, D. E.; Stubbe, J.; Kozarich, J. W.; Turner, C. J. *J. Am. Chem. Soc.* **1994**, *116*, 10843−10844. Manderville, R. A.; Ellena, J. F.; Hecht, S. M. *J. Am. Chem. Soc.* **1994**, *116*, 10851−10852.

(2) Bayer, A.; Freund, S.; Jung, G. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1336−1339. Bayer, A.; Freund, S.; Jung, G. *Eur. J. Biochem.* **1995**, *234*, 414−426.

(3) Vizan, J. L.; Hernandez-Chico, C.; Castillo, I.; Moreno, F. *EMBO J.* **1991**, *10*, 467−476.

(4) Li Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274*, 1188−1193. Stubbe, J. *Science* **1996**, *274*, 1152−1153.

(5) (a) Sinha Roy, R.; Belshaw, P. J.; Walsh, C. T. *Biochemistry* **1998**, *37*, 4125−4136. (b) Belshaw, P., Sinha Roy, R., Kelleher, N. L., Walsh, C. T. *Chem. Biol.* **1998**, *5*, 373−384.

(6) Madison, L. L.; Vivas, E. I.; Li, Y. M.; Walsh, C. T.; Kolter, R. *Mol. Microbiol.* **1997**, *23*, 161−168.

(7) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom. Rev.* **1990**, *9*, 37−70.

(8) (a) Senko, M. W.; Hendrickson, C. L.; Pasa-Tolic, L.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1824−1828. (b) Kofel, P.; Allemann, M.; Kellerhals, H. P.; Wancek, K. P. *Int. J. Mass Spectrom. Ion Processes* **1989**, *87*, 237−247.

(9) McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379−386.

(10) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

(11) The mass difference (in units of 1.0034 Da) between the most abundant isotopic peak and the monoisotopic peak is denoted in italics after each M_r value. Spectra were calibrated externally using bovine ubiquitin 8564.64-5 Da.

(12) Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. *J. Am. Chem. Soc.* **1985**, *107*, 7893.

(13) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809−2815.

S0002-7863(98)02209-4 CCC: $15.00
Published on Web 09/05/1998

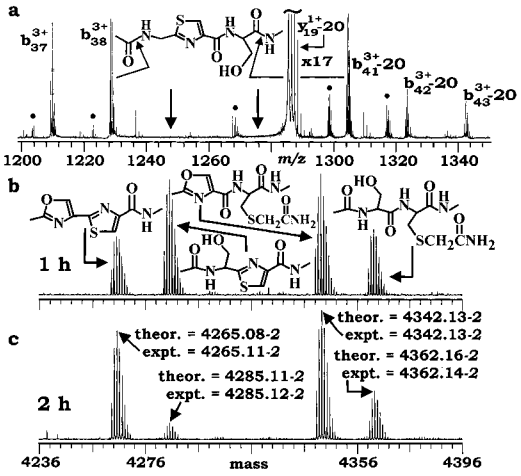


Figure 2. (a) Partial MS/MS spectrum (*m*/*z* 1200−1350) from photodissociation of $(M\text{-}20)^{4+}$ ions of the $Gly_{39}Cys_{40}Ser_{41}$ substrate; vertical arrows, theoretical positions of b_{39}^{3+} and $(b_{40}\text{-}20)^{3+}$ fragment ions; dots, H_2O loss peaks; 400 scans, 9.4 T. ESI/FT mass spectra (*x*-axis converted to mass) of the $Gly_{39}Ser_{40}Cys_{41}$ substrate after microcin synthetase processing for 1 h (b) and 2 h (c), with subsequent iodoacetamide treatment; 16 scans, 4.7 T.

arrows) while the b_{37}, b_{38}, $(b_{41}\text{-}20)^{3+}$, $(b_{42}\text{-}20)^{3+}$, and $(b_{43}\text{-}20)^{3+}$ ions are quite clear. Thus the monothiazole prevents fragmentation not only at the amide bond deleted by ring formation but also at the conjugated amide bond C-terminal to the ring. Another regioisomeric single site substrate, GGC, gave independent corroboration of this MS/MS effect, as in that case the b_{40}^{3+} and $(b_{41}\text{-}20)^{3+}$ ions were completely suppressed (>99% confidence) in the MS/MS spectrum of $(GGC\text{-}20)^{4+}$ ions (not shown). This heterocycle fingerprint now can be used analytically.

The regioisomeric double mutant $Gly_{39}Cys_{40}Ser_{41}$ (GCS), with Ser and Cys positionally interchanged from wild type McbA, has two potential sites for ring formation but has been observed to stall entirely at a single ring intermediate;[5b] MS/MS of these $(GCS\text{-}20)^{4+}$ ions shows an absence of the b_{39}^{3+} and $(b_{40}\text{-}20)^{3+}$ fragment ions (Figure 2a), which are observed for the sample before enzymatic incubation (not shown). Thus the GCS-20 species is at least 97% thiazole; any oxazole-containing species would give rise to a b_{39} ion at ∼1248 *m*/*z* (Figure 2a, vertical arrow). The thiolate anion ($Cys\ S^-$) is both more accessible than the alkoxide ($Ser\ O^-$) (pK_a 8 vs 13) and more nucleophilic, so that exclusive thiazole formation is the chemoselectivity. In addition, there may be a contributing regioselectivity programmed partially by the substrate local sequence, with cyclization to the amide carbonyl of glycine (no β-carbon) kinetically favored. The lack of any (<0.1%) bisheterocycle formation for this GCS substrate indicates that microcin synthetase in this sequence context cannot facilitate ring formation involving hydroxyl attack at an amide carbonyl that is in conjugation with a thiazole ring.

To equalize side chain reactivity and assess synthetase regioselectivity, the $Gly_{39}Cys_{40}Cys_{41}$ (GCC) and $Gly_{39}Ser_{40}Ser_{41}$ (GSS) mutant substrates were employed. The observed M_r values of the GCC- and the GSS-substrate were within 11 ppm of those expected, and MS/MS of the 4+ ions of each gave the full b_{37}^{3+}−b_{43}^{3+} fragment ion series (not shown). The substrates were enzymatically processed to ∼95% M-20 species in 8 h, with progression to M-40 bisheterocycles detectable (∼5%) but very slow. In the MS/MS spectra of both the $(GCC\text{-}20)^{4+}$ and the $(GSS\text{-}20)^{4+}$ (Supporting Information), the b_{39}^{3+} and the $(b_{40}\text{-}20)^{3+}$ ions were not present, indicating that the intermediates are >90% and >92% N-terminal heterocycle, respectively. Even with equal intrinsic reactivities of two competing nucleophiles, the enzyme cyclizes mostly the N-terminal site, suggesting either that this first ring formation site is in register with a N → C directionality of processing and/or the presence of the upstream, sterically permissive glycine has a significant effect on the cyclization rate. However, even with a thiolate as the intramolecular nucleophile, the enzyme does not catalyze efficient ring formation involving attack at an amide bond conjugated to a heterocyclic ring (<5% bithiazole in 8 h); therefore, a preferred N → C processing of bisheterocyclic sites appears unlikely. It may be that the C-terminal heterocycle (position 41) forms in the GCC- and GSS-substrates to a low level (<5%) and only this M-20 species is processed to the low level of observed M-40. This leads to a synthetase model wherein the M-20 intermediate with the N-terminal ring formed (path a, eq 1) is a kinetic "bottleneck" and this view is corroborated by the data below.

The wild type $Gly_{39}Ser_{40}Cys_{41}$ (GSC) sequence is the only substrate of this set to be processed to majority bisheterocyclic product. Its measured M_r value of 4305.16-*2* Da is 5 ppm away from that expected (4305.14-*2* Da), and MS/MS again yielded the diagnostic b_{37}−b_{43} ion series (not shown). MS/MS experiments of the $(GSC\text{-}20)^{4+}$ ions from reactions quenched at 1, 2, and 8 h showed that the b_{39} and b_{40} ions were absent in the 2 and 8 h data (indicating oxazole) but only the b_{40} fragment ion was completely suppressed in the 1 h data (not shown). This suggested that a mixture of oxazole (**I**) and thiazole (**II**) was present at early processing times. In principle, the oxazole/thiazole ratio could be determined by the $b_{41}\text{-}20/b_{39}$ ratio. However, iodoacetamide alkylation was used to separate **I** from **II** by adding 57.02 Da only to **I** with a remaining free cysteine thiol. For the partially processed, alkylated samples of GSC substrate at 1 h (Figure 2b), 2 h (Figure 2c), and 8 h (not shown), two conclusions can be drawn. First, both oxazole and thiazole monoheterocycles are formed (eq 1a and eq 1b followed) at a ratio of path a/b of about 0.6/1. Thus, the synthetase can equalize the effective nucleophilicity of Ser-OH and Cys-SH for cyclization on the upstream amide carbonyl. Second, the oxazole (**I**) apparently stalls at the M-20 stage while the thiazole (**II**) is processed on (path 1d) to M-40 bisheterocycle in ∼2 h. By 8 h, a 25% progression of oxazole to bisheterocycle was measured, indicating a path c/d ratio of ∼0.06/1. The enzyme has to balance chemoselectivity (nucleophilicity) and regioselectivity both in the first heterocyclization and second heterocyclization where β-substituents (Gly vs Cys/Ser vs heterocycle) and electrophilicity of the amide carbonyl differ. Thus, within this first bisheterocyclic site of McbA, processing to oxazole-thiazole occurs mainly with a C → N directionality.

Understanding the substrate requirements for efficient bisheterocycle production is critical in efforts to use the enzyme for combinatorial biosynthesis to generate combinations of tandem 4,2-heterocyclic systems for selection of new agents with biological activity. Future determination of regioselectivity of the synthetase toward substrates with up to eight sites of heterocycle formation will include MS/MS characterization of a second bisheterocyclization site, $Gly_{54}Cys_{55}Ser_{56}$ to see if indeed initial oxazole formation is the dominant kinetic pathway to the thiazole-oxazole pair.

Acknowledgment. We thank Jill Milne and Ranabir Sinha Roy for experimental assistance and helpful discussions. FTMS access and support were facilitated by Chris Hendrickson, Mark Emmett, Jon Quinn, and Alan Marshall (9.4 Tesla; National High-Field FT-ICR MS Facility, NHMFL, NSF CHE-94-13008), and Gary Kruppa and Frank Laukien (4.7 Tesla; Bruker Daltonics). This research was supported by the National Institutes of Health (C.T.W., GM 20011; N.L.K., postdoctoral fellowship 1F32AI10087-01). P.J.B. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research, and N.L.K. also received support from the Armenise Foundation.

Supporting Information Available: Partial MS/MS spectrum from dissociation of $(GCC\text{-}20)^{4+}$ and $(GSS\text{-}20)^{4+}$ ions (2 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA9822097